Hydrol. Earth Syst. Sci. Discuss., 12, 10739–10787, 2015
www.hydrol-earth-syst-sci-discuss.net/12/10739/2015/
doi:10.5194/hessd-12-10739-2015

Hydrology and Earth System Sciences
Open Access
Discussions

The WACMOS-ET project – Part 1: Tower-scale evaluation of four remote sensing-based evapotranspiration algorithms

D. Michel[1], C. Jiménez[2,3], D. G. Miralles[4,5], M. Jung[6], M. Hirschi[1], A. Ershadi[7], B. Martens[5], M. F. McCabe[7], J. B. Fisher[8], Q. Mu[9], S. I. Seneviratne[1], E. F. Wood[10], and D. Fernández-Prieto[11]

[1]Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland
[2]Estellus, Paris, France
[3]LERMA, Paris Observatory, Paris, France
[4]Department of Earth Sciences, VU University Amsterdam, Amsterdam, the Netherlands
[5]Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium
[6]Max Planck Institute for Biogeochemistry, Jena, Germany
[7]Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
[8]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA

[9]Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, Montana, USA
[10]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA
[11]ESRIN, European Space Agency, Frascati, Italy

Received: 2 October 2015 – Accepted: 5 October 2015 – Published: 20 October 2015

Correspondence to: D. Michel (dominik.michel@env.ethz.ch)

Published by Copernicus Publications on behalf of the European Geosciences Union.

HESSD
12, 10739–10787, 2015

Abstract

The WACMOS-ET project has compiled a forcing data set covering the period 2005–2007 that aims to maximize the exploitation of European Earth Observations data sets for evapotranspiration (ET) estimation. The data set was used to run 4 established ET algorithms: the Priestley–Taylor Jet Propulsion Laboratory model (PT-JPL), the Penman–Monteith algorithm from the MODIS evaporation product (PM-MOD), the Surface Energy Balance System (SEBS) and the Global Land Evaporation Amsterdam Model (GLEAM). In addition, in-situ meteorological data from 24 FLUXNET towers was used to force the models, with results from both forcing sets compared to tower-based flux observations. Model performance was assessed across several time scales using both sub-daily and daily forcings. The PT-JPL model and GLEAM provide the best performance for both satellite- and tower-based forcing as well as for the considered temporal resolutions. Simulations using the PM-MOD were mostly underestimated, while the SEBS performance was characterized by a systematic overestimation. In general, all four algorithms produce the best results in wet and moderately wet climate regimes. In dry regimes, the correlation and the absolute agreement to the reference tower ET observations were consistently lower. While ET derived with in situ forcing data agrees best with the tower measurements ($R^2 = 0.67$), the agreement of the satellite-based ET estimates is only marginally lower ($R^2 = 0.58$). Results also show similar model performance at daily and sub-daily (3-hourly) resolutions. Overall, our validation experiments against in situ measurements indicate that there is no single best-performing algorithm across all biome and forcing types. An extension of the evaluation to a larger selection of 85 towers (model inputs re-sampled to a common grid to facilitate global estimates) confirmed the original findings.

HESSD
12, 10739–10787, 2015

1 Introduction

Research on climate variability and the development of predictive capabilities relies largely on globally available reference data time series of the various components of the energy and water cycles. Turbulent fluxes of sensible and latent heat determine the development of the planetary boundary layer and thus govern the interactions between the Earth surface and the atmosphere. Evapotranspiration (ET) represents the time-integrated flux of latent heat and is an essential component of the energy and water cycle, playing a key role in meteorology and climate as well as agriculture (see, e.g. Ershadi et al., 2014).

Historically, there has been a lack of reliable estimates of turbulent fluxes, since the partitioning of the available energy at the Earth's surface into these fluxes is complex and characterized by large spatial and temporal variability. Also, these components of the energy balance cannot be monitored directly on a global scale by remote sensing techniques. Thus, efforts to produce satellite-based estimates typically involve combining multi-sensor data sets with predictive formulation of varying complexity, ranging from relatively simple empirical formulations to more complex modelling approaches (see e.g. Courault et al. (2005), Kalma et al. (2008) and Wang and Dickinson (2012) for comprehensive reviews). In recent years, such efforts have generated global ET products (Mu et al., 2007; Fisher et al., 2008; Jung et al., 2010; Zhang et al., 2010; Vinukollu et al., 2011; Miralles et al., 2011a) that have typically been evaluated by comparing them individually to in situ data or by inter-comparing them against other existing global heat flux estimates. For example, within the LandFlux-EVAL initiative of the Global Energy and Water cycle Exchanges (GEWEX) Data and Assessments Panels (GDAP), Jiménez et al. (2011) investigated 3 years (1993–1995) of global sensible and latent fluxes from a selection of 12 products, including satellite-based estimates, atmospheric reanalyses, and offline land surface model simulations, while Mueller et al. (2011) examined a total of 30 observation-based ET estimates from similar sources over the longer period of 1989–1995, while also providing a comparison

HESSD
12, 10739–10787, 2015

with global climate model simulations contributing to the Intergovernmental Panel on Climate Change (IPCC) 4th assessment report. More recently, Mueller et al. (2013) extended the previous Landflux-EVAL evaluations and presented two monthly global ET synthesis products, merged from individual data sets spanning the periods 1989–1995 and 1989–2005.

The GEWEX-LandFlux initiative is currently working towards producing an observation-based data set of heat fluxes that can be used together with related GDAP products to enable a joint analysis of the water and energy cycles (Jimenez et al., 2012). To contribute towards that goal, the European Space Agency (ESA) has conducted the Water Cycle Multi-mission Observation Strategy (WACMOS) EvapoTranspiration project (WACMOS-ET), aiming at the identification of appropriate algorithms to develop regional and global ET products. WACMOS-ET efforts have also included the compilation of a multi-sensor data set to run the ET methodologies for a 3 year period (2005–2007) that aimed at maximizing the use of European Earth Observation assets.

In WACMOS-ET, the methodologies by Su (2002) (Surface Energy Balance System, hereafter referred to as SEBS), Mu et al. (2011) (Penman–Monteith algorithm from the MODIS evaporation product, PM-MOD), Fisher et al. (2008) (Priestley–Taylor Jet Propulsion Laboratory model, PT-JPL), and Miralles et al. (2011a, b) (Global Land Evaporation Amsterdam Model, GLEAM) were selected to produce ET estimates at different temporal and spatial scales. The same algorithms have also been examined at a selection of different tower sites in a recent paper by McCabe et al. (2015) in preparation for the GEWEX-LandFlux product. In McCabe et al. (2015) the algorithms are run at 3-hourly time steps with both point-scale inputs (from tower meteorological observations) and gridded inputs (from the GEWEX-LandFlux global forcing data set) over a longer time period. Here, the ET algorithms are run with the WACMOS-ET forcings (see Sect. 2.2) and the analyses of model performance are extended to evaluate different time scales (3-hourly and daily) and time integrations (nighttime, daytime, and full-day). In a companion paper, Miralles et al. (2015) present the second part of

HESSD
12, 10739–10787, 2015

the WACMOS-ET study, in which PT-JPL, GLEAM, and PM-MOD are evaluated at the global scale.

The paper is structured as follows. First, the WACMOS-ET input data set is described in detail, together with the tower flux data used for driving and evaluating the ET models. This is followed by an evaluation of ET model performance at the tower scale using the tower eddy-covariance fluxes as the reference data set. The model evaluation is first performed over a selection of 24 stations covering 9 biomes in three continents (Europe, North America and Australia), in which models are run based on in situ and remote sensing forcing. Then the validation is extended to embrace a larger sample of 85 towers, in which models are driven only by satellite data re-sampled to a common grid. Finally, the main findings of our model evaluation at pixel scale are summarized.

2 Methods and data

Here the ET methodologies comprising the WACMOS-ET effort together with the input data sets that have been compiled to run the models and evaluate the ET estimates are presented. A summary of the data sets and the model-specific forcing requirements is provided in Table 1.

2.1 ET models

2.1.1 SEBS

SEBS (Su, 2002) is a one-source energy balance algorithm that is arguably one of the most widely used energy balance approaches to derive turbulent fluxes. The SEBS model calculates the sensible heat flux (H) based on the Monin and Obukhov theory (Monin and Obukhov, 1954):

HESSD
12, 10739–10787, 2015

$$\theta_0 - \theta_a = \frac{H}{k u_* \rho c_p}\left[\ln\left(\frac{z-d_0}{z_{0h}}\right) - \Gamma_h\left(\frac{z-d_0}{L}\right) + \Gamma_h\left(\frac{z_{0h}}{L}\right)\right] \quad (1)$$

$$u = \frac{u_*}{k}\left[\ln\left(\frac{z-d_0}{z_{0m}}\right) - \Gamma_m\left(\frac{z-d_0}{L}\right) + \Gamma_m\left(\frac{z_{0h}}{L}\right)\right] \quad (2)$$

$$L = -\frac{\rho c_p u_*^3 \theta_v}{kgH}, \quad (3)$$

where u is the wind speed, u_* is the friction velocity, k is the von Kármán constant, z is the height above the surface, d_0 is the zero plane displacement height, z_{0m} and z_{0h} are the roughness heights for momentum and heat transfer, Γ_m and Γ_h are the stability correction functions for momentum and sensible heat transfer, respectively. L refers to the Obukhov length, ρ is the air density, θ_0 is potential land surface temperature and θ_a is the potential air temperature at height z, g is the gravity acceleration and θ_v is the potential virtual air temperature at level z. When the suitable data are available, the only unknowns are H, u_* and L. This allows the calculation of H and the further estimation of ET based on closing the energy balance at the surface, i.e. ET is estimated as the difference between net radiation (R_n) and the sum of the calculated H and ground flux (G).

Additionally, in order to constrain H estimates, two limiting cases are considered that set upper and lower bounds for the evaporative fraction. Under very dry conditions, ET becomes zero and H is at its maximum, set by $R_n - G$. Under wet conditions, ET occurs at potential rates and therefore H is at its minimum. In this wet case, H is calculated via reverse application of the Penman–Monteith equation (see Sect. 2.1.2) assuming that the surface resistance is zero.

SEBS has been extensively validated against tower measurements and has proved to estimate realistic evaporation rates at a variety of scales, ranging from local to regional (Jia et al., 2003; Su et al., 2005; McCabe and Wood, 2006). As an example, Chen et al. (2015) recently reported an average correlation of ~0.8 and root mean

HESSD
12, 10739–10787, 2015

square difference (RMSD) of 0.7 mm day^{-1} against eddy-covariance measurements in a validation of monthly SEBS ET aggregates over China.

As a one-source energy model, SEBS does not separate the contributing components of ET (i.e. transpiration, interception loss, bare-soil evaporation), unlike the other models studied in WACMOS-ET, which provide independent estimates of these vapour sources. Although not examined here, we note that two-source energy balance models can also treat the soil and vegetation components separately (e.g. Kustas and Norman, 2000; Anderson et al., 2007), but have had limited application at the global scales.

2.1.2 PM-MOD

PM-MOD (Mu et al., 2011) is based on the Penman–Monteith equation (Monteith, 1965). It estimates ET as the sum of interception loss (I), transpiration (ET_t) and evaporation from the soil (ET_s). The interception loss is modelled as:

$$I = f_{wet} f_c \frac{\Delta(R_n - G) + \rho c_p \mathrm{VPD}/r_a^{wc}}{\lambda\left(\Delta + \gamma \frac{r_s^{wc}}{r_a^{wc}}\right)}, \quad (4)$$

where Δ is the the slope of the curve relating saturated water vapor pressure to temperature, VPD is the vapor pressure deficit, γ is the psychrometric constant, f_c is the canopy fraction, f_{wet} is the wet cover fraction based on the derivation by Fisher et al. (2008) (see Eq. 9 in Sect. 2.1.3), and r_a^{wc} and r_s^{wc} are aerodynamic and surface resistances against evaporation of intercepted water (calculated as a function of air temperature and Leaf Area Index, LAI).

Canopy transpiration is estimated as:

$$ET_t = (1 - f_{wet}) f_c \frac{\Delta(R_n - G) + \rho c_p \mathrm{VPD}/r_a^{t}}{\lambda\left(\Delta + \gamma \frac{r_s^{t}}{r_a^{t}}\right)}, \quad (5)$$

HESSD
12, 10739–10787, 2015

where r_a^t and r_s^t are the aerodynamic and surface resistances against transpiration. r_a^t is determined in a similar way to r_a^{wc} and r_s^t is a function of stomatal conductance, biome-constant values of cuticular conductance and canopy boundary-layer conductance. The values of stomatal conductance are a function of air temperature, VPD, and LAI.

Evaporation from the soil surface is the sum of evaporation from wet soil and evaporation from saturated soil, which are both calculated separately based on Eq. (7) with specific values of aerodynamic and surface resistances for bare soils and a soil moisture constraint (f_{sm}) depending on relative humidity (taken from Fisher et al. (2008), see Sect. 2.1.3).

The Mu et al. (2011) daily ET estimates have been previously validated against EC measurements from 46 FLUXNET towers in North America, reporting for the daily estimates an average RMSD of $\sim 0.9\,\mathrm{mm\,day^{-1}}$, and a ~ 0.6 average correlation coefficient.

2.1.3 PT-JPL

PT-JPL (Fisher et al., 2008) is based upon the Priestley and Taylor equation (Priestley and Taylor, 1972). As in PM-MOD, ET is estimated as the sum of interception loss I, transpiration ET_t, and evaporation from the soil ET_s. The driving equations in the model are:

$$ET_t = (1 - f_{wet}) f_g f_T f_M \alpha \frac{\Delta}{\Delta + \gamma} R_n^c \quad (6)$$

$$ET_s = f_{wet} + f_{sm}(1 - f_{wet}) \alpha \frac{\Delta}{\Delta + \gamma} (R_n^s - G) \quad (7)$$

$$I = f_{wet} \alpha \frac{\Delta}{\Delta + \gamma} R_n^c, \quad (8)$$

where α is known as the PT coefficient and is considered here as a constant value (1.26) (Priestley and Taylor, 1972) that aims to summarize the atmospheric term in the

HESSD
12, 10739–10787, 2015

Penman–Monteith equation (Eq. 5), λ is the latent heat of vaporization and f_{wet}, f_g, f_M, f_{sm} and f_T are eco-physiological constraint functions with values between 0–1 referred to as f functions. The f functions are given by:

$$f_{wet} = RH^4 \quad (9)$$

$$f_g = fAPAR/fIPAR \quad (10)$$

$$f_M = fAPAR/fAPAR_{max} \quad (11)$$

$$f_{sm} = RH^{VPD} \quad (12)$$

$$f_T = e^{-\left(\frac{T_s - T_{opt}}{T_{opt}}\right)^2}, \quad (13)$$

where f_{wet} is the relative surface wetness, f_g is green canopy fraction, fAPAR (fIPAR) is the Fraction of Absorbed (Intercepted) Photosynthetically Active Radiation, f_M is a plant moisture constraint, fAPAR$_{max}$ is the maximum of fAPAR, f_{sm} is a soil moisture constraint, f_T is a plant temperature constraint and T_{opt} is the optimum plant growth temperature, estimated as the air temperature at the time of peak canopy activity when the highest fAPAR and minimum VPD occur.

Using this methodology, monthly estimates of ET were tested against EC measurements from 16 FLUXNET towers worldwide (Baldocchi et al., 2001) with a reported average RMSD of $\sim 0.4\,mm\,day^{-1}$, and a ~ 0.9 average correlation coefficient (Fisher et al., 2008). Notice that unlike the above statistics reported for SEBS and PM-MOD, these numbers come from the model run with the tower meteorology, instead of global forcings.

2.1.4 GLEAM

GLEAM (Miralles et al., 2011a, b) calculates ET via the Priestley and Taylor (PT) equation, a soil moisture-stress computation and a Gash analytical model of rainfall interception loss (Gash, 1979). In the absence of snow, evaporation from land is calculated as:

HESSD
12, 10739–10787, 2015

$$ET = ET_{tc} + ET_{sc} + ET_s + \beta I \quad (14)$$

where ET_{tc} is transpiration from tall canopy, ET_{sc} is transpiration from short vegetation, ET_s is soil evaporation and I is tall canopy interception loss. β is a constant used to account for the times in which vegetation is wet thus transpiration occurs at lower rates ($\beta = 0.93$) (Gash and Stewart, 1977).

The first three terms in Eq. (14) are derived using the Priestley and Taylor equation, so ET becomes:

$$ET = \frac{\Delta\left[f_{tc}S_{tc}\alpha_{tc}(R_n^{tc} - G_{tc}) + f_{sc}S_{sc}\alpha_{sc}(R_n^{sc} - G_{sc}) + f_sS_s\alpha_s(R_n^s - G_s)\right]}{\lambda(\Delta + \gamma)} + \beta I, \quad (15)$$

where the subscripts tc, sc and s correspond to tall vegetation, short vegetation and bare soil (respectively), and the fraction of each of these three cover types per pixel is represented by f. Different cover types have different values of α and parameterizations of G; additionally, R_n is distributed within the cover fractions using average albedo ratios from literature. S represents the evaporative stress due to soil moisture deficit and vegetation phenology. Soil moisture deficit is estimated using a multi-layer running water balance to describe the infiltration of observed precipitation through the vertical soil profile. Microwave soil moisture observations are assimilated into the soil profile (Martens et al., 2015). In vegetated land covers, phenology effects on ET are based on microwave observations of vegetation optical depth, used as a proxy of vegetation water content.

I is independently derived using a Gash analytical model (Gash, 1979), in which a running water balance for canopies and trunks is driven by observations of precipitation. The derivation of the parameters, global implementation and validation of this I model is described in Miralles et al. (2010). For regions covered by ice and snow, sublimation is calculated based on a PT equation parameterized for ice and super-cooled waters (Murphy and Koop, 2005).

HESSD
12, 10739–10787, 2015

The ET estimates from GLEAM have been validated against eddy covariance towers worldwide; reported average correlations are 0.83 and 0.90 for daily and monthly estimates, respectively, with an average RMSD of $\sim 0.3\,\mathrm{mm\,day^{-1}}$, based on a sample of 43 towers (Miralles et al., 2011a), and correlations of 0.71–0.75 and 0.81–0.86 for daily and monthly estimates, respectively, based on a sample of 163 towers and different satellite products as forcing (Miralles et al., 2015).

2.2 Model inputs

2.2.1 Surface meteorology

The European Centre for Medium-range Weather Forecasts (ECMWF) Re-Analysis-Interim (ERA-Interim) (Dee et al., 2011) was selected to provide the near-surface meteorology every 3 h at a spatial resolution of ~ 75 km. The use of reanalysis data is necessary as satellite observations are generally unable to retrieve needed surface variables, such as temperature, humidity and wind-speed, with sufficient accuracy or at a suitable sub-daily temporal resolution. Although products of near-surface air temperature and humidity derived from satellite sounders exist (Ferguson and Wood, 2010), atmospheric reanalyses have the advantage of providing regular sub-daily estimates for all-weather conditions. ERA-Interim is also used in the derivation of the land surface temperature product (see Sect. 2.2.2), to assure inter-product consistency between air and surface temperatures. In terms of accuracy, ERA-Interim data have been evaluated through comparison with other reanalyses and weather stations over specific areas, detecting a good general performance (e.g. Mooney et al., 2011; Szczypta et al., 2011).

2.2.2 Land surface temperature

Land surface temperature (LST) estimates have been internally generated by the project from level 1 radiances from the Advanced Along-Track Scanning Radiometer

HESSD
12, 10739–10787, 2015

(AATSR) onboard ESA's EnviSAT polar-orbiting satellite, from Multi-functional Transport Satellites (MT-SAT) 2 (over Australia), Meteosat Second Generation (MSG) 2 and Geostationary Operational Environmental Satellite (GOES) 12. The data sets are provided over a sinusoidal grid with 1 km resolution for AATSR at the two satellite overpasses per day (~ 10:00 LT descending node) and 5 km for the remaining sensors (1-hourly estimates for MSG and MTSAT, 3-hourly for GOES). Ancillary atmospheric information for the inversion of the L1 radiances comes from ERA-Interim. Estimates of surface emissivity are taken from the Global Infrared Land Surface Emissivity UW-Madison Baseline Fit Emissivity Database developed by Seemann et al. (2008).

2.2.3 Surface radiation

The National Aeronautics and Space Administration (NASA)/GEWEX Surface Radiation Budget (SRB) satellite product version 3.1 (Stackhouse et al., 2004) is used to provide the surface net radiation input to the ET models. The SRB product is used by a large number of global ET algorithms to characterize the radiation at the surface, given its relatively long data record and sub-daily temporal resolution. SRB data sets include global 3-hourly averages of surface and top-of-atmosphere longwave and shortwave radiative parameters on a ~ 100 km grid.

2.2.4 LAI and *f*APAR

To characterize the vegetation state using visible and near infrared wavebands, estimates of LAI and *f*APAR have been derived by applying the Joint Research Centre (JRC) two-stream inversion package (hereafter TIP) (Pinty et al., 2007, 2011a, b) on the ESA GlobAlbedo bihemispherical reflectances. Here, LAI is defined as the one-sided leaf area per unit ground area, and *f*APAR as the fraction of absorbed in the 400–700 nm region.

The application of the TIP LAI(/*f*APAR) with our ET models required some LAI(/*f*APAR) calibration. The TIP LAI is a one-dimension (1-D) equivalent LAI for solv-

HESSD
12, 10739–10787, 2015

ing the radiative transfer in a three-dimensional medium, and it is consistent with the fluxes from which it is derived. It is not consistent with LAI derived using a 3-D radiative scheme that allows some form of horizontal clumping (e.g. the MOderate Resolution Imaging Radiometer (MODIS) MOD15A2 LAI product). In practical terms, this means that if an ET model was constructed to use a MODIS-like LAI(/*f*APAR), a straight use of the project LAI(/*f*APAR) will result in the ET model seeing lower values than expected for those biomes where horizontal clumping is significant (e.g. for forests). While the ET dynamics might not be highly affected (there is a high degree of correlation between different LAI(/*f*APAR) estimates), the absolute values would be. As the SEBS, PM-MOD, and PT-JPL models have typically been used with the MODIS vegetation product, a rescaling between our TIP-derived LAI and *f*APAR products against the MODIS product has been undertaken. For running the models at the tower scale, a local rescaling is conducted by a linear regression between the MOD15A2 and the TIP values co-registered at each tower. For global model simulations, individual rescaling per biome/climate classification is conducted.

Figure 1 illustrates an example of both products at two towers. The station *Quebec – Eastern Boreal, Mature Black Spruce* (CA-Qfo) is located in an evergreen needleleaf forest and shows that the MOD15A2 and WACMOS-ET LAI(/*f*APAR) absolute values differ considerably. This is expected, as allowing some form of horizontal clumping (MODIS 3-D radiative transfer scheme) or not (TIP 1-D) can result in large differences in the estimated LAI(/*f*APAR) in forests. It can be seen that the local calibration of the MODIS-like product retains the dynamics of the WACMOS-ET product, while adding absolute values close to the MODIS product. The station *Brookings* (US-Bkg) is situated in a cropland area, where the effects of clumping are much less severe, and the different LAI(/*f*APAR) values are much closer.

2.2.5 Vegetation height

Vegetation height at global scale is required by SEBS. For shrubland and forest biomes the product developed by Simard et al. (2011) was used as static canopy height cover.

HESSD
12, 10739–10787, 2015

For grassland and cropland biomes, where the temporal dynamics of canopy height can be more considerable, we approximated canopy height with the method by Chen et al. (2015), with the minimum and maximum canopy height obtained from the static vegetation table of the North American Land Data Assimilation System (NLDAS).

2.2.6 Soil moisture and vegetation optical depth

A soil moisture product combining observations from active and passive microwave sensors has been developed as part of the ESA Climate Change Initiative (CCI), and is adopted here to provide the surface soil moisture data that is assimilated into GLEAM. Details on the product algorithm and evaluation can be found in Liu et al. (2011b). The data is provided on a regular grid with a resolution of 0.25° × 0.25°. The product performs well in moderately vegetated regions, but shows higher uncertainties in densely vegetated regions (as vegetation attenuates the microwave signal from the ground) and mountainous areas (due to the high surface roughness) (Liu et al., 2011b).

The retrieval of soil moisture from passive sensors discussed in Sect. 2.2.6 can be accompanied by an estimation of the vegetation optical depth (VOD). VOD can be used to account for the development of vegetation over the year as it is a good proxy of vegetation water content (Liu et al., 2015). Although most ET models traditionally use parameters derived from visible and near-infrared wavelengths, microwave VOD is used by GLEAM. Here the long term record by Liu et al. (2011a) based on the application of the Land Parameter Retrieval Model by Owe et al. (2001) is used by GLEAM.

2.2.7 Precipitation and snow

Observations of precipitation and snow water equivalent are also required by GLEAM only. Precipitation is used both to estimate the effects of soil water limitations on ET and to calculate interception loss. To run the model at the tower scale we use the Climate Prediction Center (CPC) Morphing Technique (CMORPH) (Joyce et al., 2004).

HESSD
12, 10739–10787, 2015

CMORPH transports the features of precipitation estimates derived from low orbiter satellite microwave observations using information from geostationary satellite IR data. Precipitation estimates are available every 30 min on a grid with a spacing of 8 km at the equator, although the resolution of the individual satellite-derived estimates is coarser at ~ 12 km × 15 km. The spatial coverage ranges from 60° N–60° S. To run the model globally, we use the National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR)-Land precipitation estimates (Coccia and Wood, 2015). These precipitation estimates come from the hourly CFSR output (Saha et al., 2010), but are corrected using the observation-based data sets of the Climate Prediction Center (CPC) (Xie and Arkin, 1997) and the Global Precipitation Climatology Project (GPCP) (Adler et al., 2003). Finally, snow water equivalent estimates come from ESA GlobSnow. Since GlobSnow covers the Northern Hemisphere only, data from the National Snow and Ice Data Center (NSIDC) are used in snow-covered regions of the Southern Hemisphere (Kelly et al., 2003). The product combines satellite passive microwave measurements with ground based weather station data in a data assimilation scheme (Luojus and Pulliainen, 2010). The products exist at daily resolution and a spatial resolution of 25 km.

2.3 Tower data

2.3.1 Tower selection

Model simulations are evaluated by comparison with the turbulent latent fluxes measured by the eddy covariance (EC) technique at a selection of tower sites from FLUXNET (Baldocchi et al., 2001). A first sample of towers was compiled by selecting those stations from the FLUXNET La Thuile synthesis data set, which contain latent flux measurements in the 2005–2007 period, as well as the meteorological and radiation inputs required to run the ET models at the towers locations. The 24 selected stations are described in Table 2 and their geographical location is displayed in Fig. 2. While some meteorological variables such as near-surface air temperature or humidity

HESSD
12, 10739–10787, 2015

are measured at nearly all towers, other inputs such as the surface net radiation or the ground heat flux are measured at only a few towers. Some stations that were very close to the shore or in places with regular flooding were discarded. The final selection of 24 towers represents a significant number of biomes and a reasonable sample of dry and wet climate regimes.

In a later section, by removing the constraint of requiring local measurements of all the model inputs, the first selection of 24 towers is extended to a total of 85 stations. This second selection is used to evaluate model performance when the models are run with the satellite data used for the global runs.

2.3.2 In situ surface energy balance

While, in principle, the surface energy balance should close at the tower, this is rarely the case: a lack of closure in the surface energy balance of about 10–30 % is commonly found when comparing the EC measurements against the energy balance residual (ER) term, i.e. the difference between net radiation and the sum of the sensible, latent and ground fluxes (e.g. Foken et al., 2006). Consequently, throughout the paper the model evaluation is discussed by comparing against both the EC measurements and the in situ ER estimates.

2.3.3 In situ LST

To run SEBS, the broadband longwave radiometer measurements need to be converted into LST estimates. This is done by inverting the equation relating the upwelling spectral radiance measured by the radiometer and the LST. Broadband emissivity is required and it is estimated from the MODIS-based Global Infrared Land Surface Emissivity Database (Seemann et al., 2008) operated by the Cooperative Institute for Meteorological Satellite Studies (CIMSS). The estimates are calculated by following the approach suggested by Wang et al. (2005) using a linear combination of narrowband emissivities at 8.5, 11, and 12 µm.

HESSD
12, 10739–10787, 2015

2.3.4 In situ vegetation height

SEBS also requires vegetation height to derive the surface roughness values. In most cases a mean annual value can be obtained from the tower metadata and this value is adopted here as vegetation height at the tower. However, a clear limitation in this assumption is that it does not include dynamic changes in vegetation height over time. As discussed in Sect. 2.2, the importance of neglecting the temporal variability in height is biome-dependent; for instance, in forests the mean vegetation height is typically more constant than in, e.g. croplands, where the changes derived from agricultural practices can be large.

2.4 ET experiments

2.4.1 Evaluation times

The model performance is investigated at sub-daily and daily time scales. The tower data is available at 1/2 h intervals and has been time-integrated to 3-hourly in order to run the ET models at that sub-daily resolution. The satellite data has been time-matched to the 3-hourly or daily resolutions from their native resolution in different ways (see below), depending on the type of data and original resolution. The 3-hourly inputs were then aggregated to daily values in order to run the models with tower-based daily inputs. The tower data record is not always time-continuous, as in some instances there are gaps in the record. This is not a problem for the PM-MOD, PT-JPL and SEBS model, because the ET estimates depends only on the instantaneous atmospheric/surface state. When inputs to the models and/or ET for the evaluation are missing, those three models are not run. Conversely, GLEAM requires continuous data records to update the soil moisture state variable. To facilitate running GLEAM with tower inputs, the tower measurements are gap-filled with the corresponding pixel data (see Sect. 2.2). ET estimates from those periods are removed after the runs, so as

HESSD
12, 10739–10787, 2015

before, only the time steps where tower forcing data are available are used for model evaluation.

The models are validated against the tower ET only under dry (non-raining) conditions, as EC gas analyzers are not reliable during rain events due to disturbance of the infrared signal by droplets on the sensor (Burba et al., 2010; Hirschi et al., 2015). Therefore, any days with precipitation as indicated by the tower or satellite precipitation are removed from the validation, and the interception component from PM-MOD, PT-JPL, and GLEAM is not considered in the validations.

2.4.2 Nighttime ET

Only PM-MOD and GLEAM specifically deal with nighttime evaporation. Nevertheless, nighttime values are required from all models to integrate the 3-hourly ET estimates to daily values. For SEBS and PT-JPL negative nighttime estimates are set to zero to allow the daily integration for those models. To separate day and night, daylight times are identified by calculating the solar zenith angle. Time intervals, where the cosine of the zenith angle is larger than 0.2, are kept as day values. This day and night separation may be less accurate than using a solar downward radiation threshold, but it allows a day-night flag for those stations without solar radiation measurements. The impact of setting ET from SEBS and PT-JPL to zero, as these models cannot specifically produce nighttime, is addressed in Sects. 3.1 and 3.2, where sub-daily periods, including daytime, are investigated.

2.4.3 ET production

The following ET estimates are generated to evaluate model performance:

– *tower-based ET:* ET generated by the 4 models using the 3-hourly or daily in situ data (surface radiation, LST, air temperature, air humidity and wind speed and precipitation), the scaled WACMOS-ET LAI/*f*APAR and gridded soil moisture and VOD data. Note that the 3-hourly tower-based ET estimates are also time-

HESSD
12, 10739–10787, 2015

integrated to daily values, so daily ET estimates exist both from the runs with daily inputs and from the integration of the 3-hourly ET estimates.

– *original-resolution satellite-based ET:* ET generated by the 4 models using the 3-hourly or daily satellite data (SRB surface radiation, ERA-Interim air temperature, humidity, wind speed, CMORPH precipitation, ESA-CCI soil moisture, scaled WACMOS-ET LAI/*f* APAR) at their original resolutions. In situ LST is still used here in order to have the same number of SEBS estimates as in the tower-based ET (cloudiness, satellite overpass time and revisiting times would have notably reduced the number of SEBS estimates if the satellite LST had been used). As for the tower-based ET, 3-hourly tower-based ET estimates are also time-integrated to daily values.

– *common-grid satellite-based ET:* ET generated by the 4 models using the 3-hourly satellite data re-sampled to a common grid. In contrast to the previous runs, the satellite data is not applied at their original resolutions, but after re-sampling them to the sinusoidal grid at ∼ 25 km adopted to produce the global model runs. Note that the CMORPH precipitation is replaced by the CFSR-Land product in order to have global coverage, and that the LST is based on the AATSR observations.

3 Results and discussion

Here we look at the model performance against the in situ measurements, when the models are run with tower-based and satellite inputs. This section is divided into the three subsections, each of them dealing with one of the three experiments introduced in the previous Sect. 2.4.3. First, the 3-hourly and daily runs based on in situ forcing at 24 FLUXNET stations (see Table 2) are investigated. In the second part we look at the model performance at the same stations using 3-hourly and daily resolution satellite forcing. Finally, the ET estimates from the run using 3-hourly common-grid satellite forcing are compared to the in situ measurements at 85 FLUXNET stations.

HESSD
12, 10739–10787, 2015

3.1 3-hourly and daily tower-based ET

The agreement of modelled evaporative fraction (EF) – defined here as $\lambda E/R_n$ using modelled λE and the net radiation from the respective forcing – against the measured EF (i.e. based on tower measurements of λE and R_n) gives an indication of the algorithm skill to model evaporative stress. Figure 3 (top panel) illustrates the agreement of modelled evaporative fraction to in situ measurements (derived using both EC and ER measurements of evapotranspiration, see Sect. 2.3.2), when models are run with tower inputs. GLEAM generally ranges between the EC and ER measurements, even at dry stations in open shrubland (OSH) and grassland (GRA) biomes (e.g. *Sardinia/Arca di Noe* (IT-Noe), *Audubon Research Ranch* (US-Aud), *Santa Rita Mesquite* (US-SRM) and *Walnut Gulch Kendall Grasslands* (US-Wkg)). Only in the evergreen needleleaf forests (ENF) does GLEAM exceed the range of in situ measurements. PT-JPL mostly agrees to the reference as well, although it presents positive biases at some dry sites, like *Wind River Crane* (US-Wrc) and IT-Noe. PM-MOD underestimates EF for most stations (but it is very close to the EC measurements at 6 stations), while SEBS is characterized by an overall overestimation (for 6 stations SEBS EF is within the tower EC-ER range). In terms of the model performance per biome type, it can be stated that models generally perform the best in croplands (CRO) and deciduous broadleaf forest (DBF); at least this is the case for PT-JPL, PM-MOD and GLEAM. SEBS seems to perform better in grassland and savanna biomes (SAV). It is, however, difficult to derive robust conclusions on the model performance as function of biome due to the low number of stations per biome type.

As the surface meteorology plays an important role in the ET production, we also compare the point-scale model performance with the gridded ERA-Interim ET data set (ERA) in Fig. 3 (top panel). ERA-Interim estimates are mostly within the range of EF measurements. The good agreement between ERA EF and the in situ measurements indicates that the ERA-Interim meteorology reliably captures the station conditions.

HESSD
12, 10739–10787, 2015

It can also be stated that the point-scale tower-forced EF derived with PT-JPL and GLEAM match the ERA-Interim product based on a ~ 75 km resolution.

A statistical assessment of the model performance is given in Fig. 4, which shows the correlation (R^2), the RMSD and the average of the bias normalized by the reference (MBD) between modelled ET and tower measurement of ET (i.e. using the EC approach). In the left column of Fig. 4 the station averages of the statistical inferences are shown in the order of measured EF, i.e. from wet to dry. In general, the correlation to in situ data is high in wet and in moderately wet biomes for most sites and for all models. This is also true for SEBS, despite its substantial overestimation of EF (see Fig. 3). However, there seems to be a distinct decrease of R^2 from wet to dry biomes for all models; this decrease in performance is lower for GLEAM and higher for PM-MOD, that presents correlations ($R^2 < 0.4$) at dry sites. PT-JPL stands out of the ensemble with the highest correlation at most sites and especially in dry conditions. In comparison to the mostly underestimated evaporative fraction derived with PM-MOD (see Fig. 3), the RMSD of PM-MOD ET corresponds to PT-JPL and GLEAM, and even produces the lowest maximum value (0.13 mm h^{-1}), followed by GLEAM (0.17 mm h^{-1}). Note that the large positive MBD values of PT-JPL and SEBS (> 200 %) may partly result from forcing ET to zero during nighttime (see Sect. 2.4.2), when tower ET is negative, and thus leading to large relative errors, even for small negative reference ET values.

In order to evaluate the impact of using EC measurements as reference (in contrast to the ER method in Fig. 4), Table 3 shows the overall average 3-hourly model performance (i.e. the average of all station statistics) using both EC and ER data as reference. Overall, the average statistics of PT-JPL and GLEAM appear more favourable than those of SEBS and PM-Mu, although the RMSD and MBD of PM-MOD and the R^2 of SEBS are in general comparable to those of GLEAM and PT-JPL. This is again to a large extent affected by the overall overestimation and underestimation by SEBS and PM-MOD, respectively. The RMSD of SEBS is significantly smaller when using the ER method as reference (0.10 mm h^{-1}) as opposed to using EC (0.13 mm h^{-1}); on the other hand, the RMSD of PM-MOD is larger against ER (0.12 mm h^{-1}) than against

HESSD
12, 10739–10787, 2015

EC (from 0.06 mm h^{-1}). Note that the transpiration resistances in PM-MOD are calibrated based on a biome-dependent annual ET derived from EC observations, which may explain the smaller RMSD and MBD when using EC as a reference. Finally, the RMSD station averages are similar against both in situ references for by PT-JPL (0.08, 0.09 mm h^{-1}) and GLEAM (0.08, 0.08 mm h^{-1}).

Here the skill of models at representing ET at specific times of the day is examined. Note that small nighttime ET values from models and measurements may produce small absolute errors and thus can improve the overall full day model performance in comparison to daytime periods, even if the relative bias is large.

The Taylor diagrams in Fig. 5 show that correlations to in situ observations (using the EC method) are higher when the entire daily cycle is considered (left panel), as opposed to considering daytime values only (left panel, top row) or nighttime values only (right panel, top row). The overall R^2 with tower forcing including all models increases from 0.54 to 0.67 from daytime to full day evaluation; this reflects the fact that the daily solar cycle leads to preferentially high values around noon and lower at night, which will increase correlations as long as the models are able to reproduce the sensitivity to radiation changes adequately. Figure 5 (right panel, top row) shows the the overall model performance for nighttime periods. Note that nighttime is identified as cases, when the cosine of the zenith angle is < 0.2. We can see that forcing negative nighttime ET values of PT-JPL and SEBS to zero (in contrast to specific negative ET produced by PM-MOD and GLEAM, see Sect. 2.4.2) does not have a substantial impact on the overall agreement to tower measurements. However, it should be noted that the uncertainty of nighttime EC measurements is high because of low turbulence. Hence, large nighttime errors can be present not only in the ET simulations but also in the EC data.

Sub-daily resolution is desirable in evaporation modelling, as it allows investigation of the underlying land-atmospheric interactions during the daily cycle of the planetary boundary layer. Given the short time scale of these interactions, one may expect that models that are able to reproduce short-term variability in ET would also be able to provide more reliable aggregates at daily time scales. Therefore, we investigate whether

HESSD
12, 10739–10787, 2015

the model performance would benefit from solving evaporation at 3-hourly resolution and aggregating these to daily, as opposed to generating the estimates with daily input directly. Figure 5 (bottom row) clearly shows that not much more skill is gained by producing daily ET based on 3-hourly input (i.e. resolved diurnal cycles in the meteorological inputs) as opposed to forcing the models with the original daily input; results are almost identical when using aggregated 3-hourly output (left panel, bottom row) or using daily forcing (right panel, bottom row). In fact, for GLEAM the correlation to the EC reference is slightly higher when daily input is used, even if the standard deviation agrees marginally worse to the reference.

Figure 6 shows the statistics of the models evaluation after forcing them with daily inputs. As expected, the general correlations become lower when daily (as opposed to 3-hourly) estimates are validated, since the daily cycle no longer plays a role on the enhancement of correlations – this was already highlighted by Table 3. Comparison of Figs. 4 and 6 shows that the decline of average R^2 from wet to dry stations is less evident at daily resolution. This can be affected by the smaller sample size when daily values are analyzed. PM-MOD and SEBS in particular correlate poorly at dry stations (also at other stations, such as the moderately-wet AU-How). PT-JPL and GLEAM perform worse (compared with the 3-hourly) at dry stations when they are run at daily resolutions. In terms of the RMSD and MBD, the results are quite similar to the 3-hourly findings, but in most cases worst performance at the daily resolution is found at dry stations. An exception is GLEAM, which shows smaller RMSD at the dry stations compared with the 3-hourly.

The change of overall MBD (against the EC reference) from using 3-hourly tower input to using daily tower input for PT-JPL is from 53.1 to 47.8 %, for PM-MOD from −6.7 to 3.8 %, for SEBS from 125.9 to 113.5 % and for GLEAM from 31.9 to 15.6 %. While the pattern of EF (Fig. 3) and MBD (Fig. 4) indicates a substantial underestimation of 3-hourly ET by PM-MOD, this underestimation is attenuated when daily input is used (−18.2 to 11.3 % against the ER reference). Note that even if we employ the term daily input, the PM-MOD model estimates day and night ET separately by using integrated

HESSD
12, 10739–10787, 2015

day and night inputs (as opposed to PT-JPL, SEBS, and GLEAM, which use daily integrated inputs), and then combines them to provide a daily value. This is how the PM-MOD model was originally used and how it is implemented in this study for daily estimation. The better agreement at daily scale thus may reflect a more appropriate use of their inputs.

The similarity of the results for different temporal resolutions underlines the robustness of the modelling processes. PT-JPL and GLEAM agree best to the in situ measurements, while SEBS yields a good correlation in comparison to the other models, yet produces the largest absolute errors due to its large overestimation. PM-MOD produces the lowest correlation but agrees rather well in terms of absolute deviations.

Table 3 summarizes the main statistics of the model evaluation for the 3-hourly and daily tower inputs.

3.2 3-hourly and daily original-resolution satellite ET

In this section we discuss the model performance using 3-hourly and daily satellite forcing with original resolution at the selected 24 FLUXNET stations. The findings are compared to the results of the tower forcing in the previous section in order to allocate model uncertainty to either the algorithms used or the common forcing.

The evaluation of 3-hourly modelled EF using satellite forcing (Fig. 3, bottom panel) shows a very similar picture of agreement to the reference compared to the results of the tower forcing. Note that the satellite EF shown here slightly differs from tower-forced EF, as the data availability of the input time series may be different at some stations. The ET overestimation by SEBS seems to be slightly emphasized when using satellite input in comparison to the tower forcing. Note that the LST used in SEBS is still obtained from the tower measurements, as discussed in Sect. 2.4.1. EF derived with PT-JPL and GLEAM still agrees well with the reference, yet GLEAM overestimates EF in dry biomes when using satellite forcing, but is more accurate at needleleaf forest sites. The good model performance of PT-JPL and GLEAM, independent of forcing type, indicates a robust performance of the models on the one hand and a reliable

HESSD
12, 10739–10787, 2015

satellite forcing – in the sense of their meteorology comparing well with the in situ tower data – on the other hand.

In Fig. 3 (bottom panel) we also compare the model performance with the gridded ERA-Interim ET data set. Note that while the tower forcing runs (top panel) are independent from ERA-Interim, the satellite runs use ERA-Interim estimates as inputs for the surface meteorology. As shown in Sect. 3.1, the ERA-Interim EF product agrees with the in situ measurements. The correlation of the models to ERA-Interim is not substantially improved with satellite input in comparison to the tower forcing, although they now use the ERA-Interim meteorology as input.

The station averages of the statistical indices R^2, RMSD and MBD of the models forced with satellite observations (Fig. 4, bottom panel) against the in situ measurements underline the previously reported high similarity of modelled ET based on tower and satellite forcing. Only the RMSD of SEBS is slightly attenuated with remotely sensed forcing. However, the algorithm is still characterized by substantial overestimation.

In the following we compare the model performance with daily satellite forcing to the model performance with daily tower forcing. In accordance with the evaluation of 3-hourly data (see Fig. 4), Fig. 6 indicates that the daily satellite-based ET products also correspond to the tower-based modelled ET. We want to highlight, however, that in contrast to the 3-hourly runs, the RMSD of SEBS substantially increases when satellite input is used. This suggests that the SEBS physical modeling captures the ET processes more accurately with the high temporal resolution inputs (3-hourly vs. daily).

Table 4 provides a summary of the main statistics of the model evaluation for the 3-hourly and daily satellite inputs.

3.3 3-hourly common-grid satellite ET

Here the ET algorithms are tested against 85 FLUXNET stations using the gridded sinusoidal (~ 25 km) satellite input (as opposed to using their original input resolutions) in order to evaluate the common-gridded global ET estimates at the tower scale. Only

HESSD
12, 10739–10787, 2015

the evaluation over the towers is discussed here, with the evaluation at the global scale discussed in the companion paper of Miralles et al. (2015). Note that the spatial mismatch between the tower fetch and the $\sim 750\,km^2$ of the gridded cells is very large, and the agreement between the tower fluxes and the modelled ET certainly depends on the tower conditions being representative for the corresponding gridded pixel. That was also the case for some of the original-resolution satellite inputs used over the 24 stations, such as the SRB radiation or the ERA-Interim meteorology. The results of the satellite runs using common-grid forcing are compared to the results using the tower and satellite inputs at the tower scale presented in Sects. 3.1 and 3.2.

The top panel of Fig. 7 shows the mean 3-hourly EF over 70 stations for PM-MOD, PT-JPL, and GLEAM. For 15 of the 85 stations the surface radiation or the ground flux was not available, hence the ER reference could not be calculated. As the gridded inputs use satellite LST from AATSR, SEBS ET is only estimated at the mid-morning AATSR overpass. The bottom panel of Fig. 7 shows the annual mid-morning evaporative fraction, this time including SEBS. Due to the 3 day revisiting time of AATSR and the lack of measurements in cloudy conditions, the number of available SEBS ET estimates reduces drastically, compared with the previous simulations using tower LST. The bottom panel of Fig. 7 shows station averages from all models only when SEBS ET is available. Thus, it is based on less data and with the number of stations reduced to 67.

The 3-hourly model performances from PM-MOD, PT-JPL, and GLEAM correspond closely to the performance over the analysis using the 24 towers and the original-resolution satellite inputs. The EF station averages produced by PT-JPL and GLEAM are very close at all locations and respond well to the hydrological and energetic conditions expected in the respective biome. The overall agreement to the range between EC and ER in situ measurements is comparable to what has previously been found in the smaller sample of stations (see Fig. 3). PM-MOD keeps underestimating ET, except for the cropland biome, where the majority of station averages match well with the reference. Concerning the mid-morning evaporative fractions, the PM-MOD, PT-

HESSD
12, 10739–10787, 2015

JPL, and GLEAM patterns are all very similar to the case with the full diurnal cycle. SEBS again tends to overestimate over a large number of stations, compared with the in situ measurements. Overall, it can be stated that the model accuracy and inter-model agreement obtained with in situ and satellite forcing at the tower scale could be reproduced with the common-grid satellite forcing.

Figure 8 summarizes the results above by displaying standard deviation, correlation and RMSD of the modelled ET shown in Fig. 7 against the EC reference. The Taylor plots highlight the fact that the variability of PT-JPL, PM-MOD and GLEAM is not substantially influenced by the low sample size for cases, when SEBS ET is available. Again, the similarity of Fig. 5 (left panel) for satellite forcing at the tower scale and Fig. 8 for gridded input data confirms the robustness of the analyses independent of tower and time sampling.

4 Conclusions

In this first part of the WACMOS-ET study, the skill of the PT-JPL, PM-MOD, SEBS and GLEAM ET algorithms have been tested at the tower scale against in situ measurements at 24 FLUXNET sites. The algorithms are forced using in situ meteorological data from these towers, covering the period 2005–2007 on three continents and across 9 different biomes – while ensuring spatial consistency between input and reference data. Additionally, the models are run for the same period with reanalysis and satellite forcing of varying spatial resolution, including ERA-Interim (surface meteorology), SRB (radiation), AATSR (LST), GlobAlbedo (LAI(/*f*APAR)), CMORPH (precipitaton) and WACMOS-CCI (soil moisture). The models were simulated with 3-hourly and daily input to assess the robustness of their performance for sub-daily and daily resolution.

Our analyses have shown that the 4 models performance is robust against changes in forcing types and temporal resolutions (i.e. the changes do not alter significantly the model behavior at the selected stations). Against the in situ 3-hourly energy resid-

HESSD
12, 10739–10787, 2015

ual estimates at the tower, the tower-based model simulations are ranked (according to station averages) as GLEAM (0.80, 0.08), PT-JPL (0.78, 0.09), SEBS (0.78, 0.10) and PM-MOD (0.55, 0.12). The first value in the brackets denotes R^2 and the second value denotes RMSD in mm h^{-1}. Against the eddy-covariance measurements however, the station averages of RMSD do not reflect the same outcome. Due to more substantial overestimation at two stations each, the RMSD of PT-JPL (0.77, 0.08) and GLEAM (0.70, 0.08) are larger than that of PM-MOD (0.58, 0.06). However, correlations keep being higher for GLEAM and PT-JPL. Thus, over our selection of towers and reference period (2005–2007) we judge GLEAM and PT-JPL as the algorithms more closely matching the in situ observations. At some stations PM-MOD and SEBS also agree well with the observations, but in general PM-MOD and SEBS performance can be characterized by under- and overestimation, respectively.

For the satellite forcing, the RMSD between the models and the reference yields very similar numbers as for tower forcing. Correlations are closer, but in most situations slightly smaller for the satellite forcings. This can be the result of discrepancies between the spatial resolution of satellite observations and tower measurements, although different inputs errors (in situ vs. satellite) may also play a role. This performance closeness between in situ- and satellite-derived can be an indication of the spatial representativeness of the tower measurements (i.e. reasonable spatial homogeneity around the tower) and the consistence of the input data set across forcing types. This is underlined by a comparison to the ~ 75 km resolved reanalysis ET product of ERA-Interim, which agrees well to the modelled ET across the different biomes.

Regarding the analysis over the 85 stations, a similar overall picture is obtained using the ~ 25 km common-grid ET prepared for the global runs. The evaluations of McCabe et al. (2015) over a different selection of towers (45 stations), a more extended period (1997–2007) and different satellite forcings (LandFlux forcings), also result in an overall similar analysis, confirming the robustness of the model performance evaluations.

Using daily input data reduces the RMSD of the models with the tower measurements, but results in slightly worse correlations. This is due to the lower variability

HESSD
12, 10739–10787, 2015

of daily values in contrast to three-hourly data (variability accentuated by the diurnal cycle). However, the consistency of the model agreements to the reference between three-hourly and daily ET estimates highlights the robustness of the integration method applied to the models. This is also underlined by the good agreements of modelled daily ET from aggregated three-hourly output data with modelled daily ET from daily input.

While GLEAM and PM-MOD can produce negative ET, PT-JPL and SEBS cannot operate under these conditions (mostly at nighttime where the flux of available energy reverses sign) and their negative values are forced to zero. This does not have a large impact on their full day performance, since these values are occurring at night, when tower ET is negative and with generally low values. Only for the relative bias is the effect significant, since the two models are consequently overestimating ET in these cases.

In terms of high and low temporal input resolution, it was found that using 3-hourly input data does not significantly increase the accuracy of the models for producing daily ET. Hence, it is sufficient to use daily input to achieve a similar result if the intended application of the ET product does not demand a reproduction of the diurnal cycle.

The conducted analyses based on in situ ET are useful to evaluate model performance, but there are some clear limitations. Our requirements for tower selection resulted in a somehow limited number of stations, so it would be desirable to extend the evaluations to larger regions in order to better cover different climate and biome conditions. Therefore, in the companion paper of Miralles et al. (2015) our analyses are extended by looking at the global spatiotemporal variability of the modelled ET, the closure of regional water budgets, and the discrete estimation of land evaporation components or sources (i.e. transpiration, interception loss and direct soil evaporation).

Acknowledgements. This study was funded by the European Space Agency (ESA) and conducted as part of the project WACMOS-ET (Contract no 4000106711/12/I-NB). D. G. Miralles acknowledges the financial support from the Netherlands Organization for Scientific Research through grant 863.14.004, and the Belgian Science Policy Office (BELSPO) in the frame of the STEREO III programme, project SAT-EX (SR/00/306). M. F. McCabe and A. Ershadi acknowledge the support of the King Abdullah University of Science and Technology. The SEBS

HESSD
12, 10739–10787, 2015

team is acknowledged for facilitating discussions concerning the implementation of their model. This work used eddy covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (U.S. Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program (DE-FG02-04ER63917 and DE-FG02-04ER63911)), AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCan), GreenGrass, KoFlux, LBA, NECC, OzFlux, TCOS-Siberia, USCCC. Data and logistical support for the station US-Wrc were provided by the US Forest Service Pacific Northwest Research Station. All WACMOS-ET forcing data and ET estimates are publicly available and can be requested through the project website (http://wacmoset.estellus.eu).

References

Adler, R., Huffman, G., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1970–Present), J. Hydrol., 4, 1147–1167, 2003. 10754

Anderson, M., Norman, J., Mecikalski, J., Otkin, J., and Kustas, W.: A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation, J. Geophys. Res., 112, D10117, doi:10.1029/2006JD007506, 2007. 10746

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K., Pilegaard, K., Schmid, H., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsyn, S.: FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, B. Am. Meteorol. Soc., 82, 2415–2434, 2001. 10748, 10754

Burba, G., McDermitt, D., Anderson, D., and Furtaw, M. D.: Novel design of an enclosed CO_2/H_2O gas analyser for eddy covariance flux measurements, Tellus, 62, 743–748, 2010. 10757

HESSD
12, 10739–10787, 2015

Chen, X., Su, Z., Ma, Y., Liu, S., Yu, Q., and Xu, Z.: Development of a 10-year (2001–2010) 0.1° data set of land-surface energy balance for mainland China, Atmos. Chem. Phys., 14, 13097–13117, doi:10.5194/acp-14-13097-2014, 2014. 10745, 10753

Coccia, G. and Wood, E. F.: CFSR-Land: a new high temporal resolution global land data assimilation product, J. Geophys. Res., in preparation, 2015. 10754

Courault, D., Seguin, B., and Olioso, A.: Review on estimation of evapotranspiration from remote sensing data: from empirical to numerical modeling approaches, Irrig. Drain. Syst., 19, 223–249, 2005. 10742

Dee, D., Uppala, M., S., Simmons, J., A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, A., M., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, M., A. C., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, J., A., Haimberger, L., Healy, B., S., Hersbach, H., Hólm, V., E., Isaksen, L., Kallberg, P., Khaler, M., Matricardi, M., McNally, P., A., Monge-Sanz, M., B., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thapaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828, 2011. 10750

Ershadi, A., McCabe, M. F., Evans, J. P., Chaney, N. W., and Wood, E. F.: Multi-site evaluation of terrestrial evaporation models using FLUXNET data, Agr. Forest Meteorol., 187, 46–61, 2014. 10742

Ferguson, C. R. and Wood, E. F.: An evaluation of satellite remote-sensing data products for land surface hydrology: Atmospheric Infrared Sounder (AIRS), J. Hydrometeorol., 11, 1234–1262, 2010. 10750

Fisher, J., Tu, K., and Baldocchi, D.: Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites, Remote Sens. Environ., 112, 901–919, 2008. 10742, 10743, 10746, 10747, 10748

Foken, T., Wimmer, F., Mauder, M., Thomas, C., and Liebethal, C.: Some aspects of the energy balance closure problem, Atmos. Chem. Phys., 6, 4395–4402, doi:10.5194/acp-6-4395-2006, 2006. 10755

Gash, J.: An analytical model of rainfall interception by forests, Q. J. Roy. Meteorol. Soc., 105, 43–45, 1979. 10748, 10749

Gash, J. and Stewart, J.: The evaporation from Thetford Forest during 1975, J. Hydrol., 35, 385–396, 1977. 10749

HESSD
12, 10739–10787, 2015

Hirschi, M., Michel, D., Lehner, I., and Seneviratne, S. I.: A site-level comparison of lysimeter and eddy-covariance flux measurements of evapotranspiration, Agr. Forest Meteorol., in review, 2015. 10757

Jia, L., van den Z. Su, B. H., Menenti, M., Moene, A., Bruin, H. A. R. D., Yrisarry, J. J. B., Ibanez, M., and Cuesta, A.: Estimation of sensible heat flux using the Surface Energy Balance System (SEBS) and ATSR measurements, Phys. Chem. Earth, 28, 75–88, 2003. 10745

Jiménez, C., Prigent, C., Mueller, B., Seneviratne, S. I., McCabe, M. F., Wood, E. F., Rossow, W. B., Balsamo, G., Betts, A. K., Dirmeyer, P. A., Fisher, J. B., Jung, M., Kanamitsu, M., Reichle, R. H., Reichstein, M., Rodell, M., Sheffield, J., Tu, K., and Wang, K.: Global intercomparison of 12 land surface heat flux estimates, J. Geophys. Res., 116, 6809–6866, D02102, doi:10.1029/2010JD014545, 2011. 10742

Jimenez, C., McCabe, M., Seneviratne, S. I., Wood, E., and Rossow, W.: 4th LandFlux Workshop, GEWEX News, 22, 17–18, 2012. 10743

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P.: CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, J. Hydrometeorol., 5, 487–503, 2004. 10753

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., and de Jeu, R.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951–954, 2010. 10742

Kalma, J., McVicar, T., and McCabe, M.: Estimating land surface evaporation: a review of methods using remotely sensed surface temperature data, Surv. Geophys., 29, 421–469, doi:10.1007/s10712-008-9037-z, 2008. 10742

Kelly, R., Chang, A., Tsang, L., and Foster, J.: A prototype AMSR-E global snow area and snow depth algorithm, IEEE T. Geosci. Remote, 41, 230–242, 2003. 10754

Kustas, W. and Norman, J.: A two-source energy balance approach using directional radiometric temperature observations for sparse canopy covered surfaces, Agron. J., 92, 847–854, 2000. 10746

Liu, Y. Y., de Jeu, R. A. M., McCabe, M., Evans, J., and van Dijk, A. I. J. M.: Global long-term passive microwave satellite-based retrievals of vegetation optical depth, Geophys. Res. Lett., 38, L18402, doi:10.1029/2011GL048684, 2011a. 10753

Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., van Dijk, A. I. J. M., McCabe, M. F., and Evans, J. P.: Developing an improved soil moisture dataset by blending

HESSD
12, 10739–10787, 2015

passive and active microwave satellite-based retrievals, Hydrol. Earth Syst. Sci., 15, 425–436, doi:10.5194/hess-15-425-2011, 2011b. 10753

Liu, Y. Y., van Dijk, A. I. J. M., de Jeu, R. A. M., Canadell, J. G., McCabe, M. F., Evans, J. P., and Wang, G.: Recent reversal in loss of global terrestrial biomass, Nat. Clim. Change, 5, 470–474, doi:10.1038/nclimate2581, 2015. 10753

Luojus, K. and Pulliainen, J.: Global Snow Monitoring for Climatic Research: Snow Water Equivalent (SWE) Product Guide, Finnish Meteorological Institute, Helsinki, Finland, 2010. 10754

Martens, B., Miralles, D. G., Lievens, H., Fernández-Prieto, D., and Verhoest, N.: Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture, Int. J. Appl. Earth Obs., in press, 2015. 10749

McCabe, M. F. and Wood, E.: Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors, Remote Sens. Environ., 105, 271–285, doi:10.1016/j.rse.2006.07.006, 2006. 10745

McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D. G., Michel, D., and Wood, E. F.: The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally-gridded forcing data, Geosci. Model Dev. Discuss., 8, 6809–6866, doi:10.5194/gmdd-8-6809-2015, 2015. 10743, 10767

Miralles, D. G., Gash, J. H., Holmes, T. R. H., de Jeu, R. A. M., and Dolman, A. J.: Global canopy interception from satellite observations, J. Geophys. Res., 115, D16122, doi:10.1029/2009JD013530, 2010. 10749

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15, 453–469, doi:10.5194/hess-15-453-2011, 2011a. 10742, 10743, 10748, 10750

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., and Dolman, A. J.: Magnitude and variability of land evaporation and its components at the global scale, Hydrol. Earth Syst. Sci., 15, 967–981, doi:10.5194/hess-15-967-2011, 2011b. 10743, 10748

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., Hirschi, M., Martens, B., Dolman, A. J., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F., and Fernandez-Prieto, D.: The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets, Hydrol. Earth Syst. Sci. Discuss., 12, 10651–10700, doi:10.5194/hessd-12-10651-2015, 2015. 10743, 10750, 10765, 10768

HESSD
12, 10739–10787, 2015

Monin, A. and Obukhov, A.: Basic laws of turbulent mixing in the surface layer of the atmosphere, Tr. Akad. Nauk SSSR Geophiz. Inst., 24, 163–187, 1954. 10744
Monteith, J.: Evaporation and environment, Sym. Soc. Exp. Biol., 19, 205–234, 1965. 10746
Mooney, P. A., Mulligan, F. J., and Fealya, R.: Comparison of ERA-40, ERA-Interim and NCEP/NCAR reanalysis data with observed surface air temperatures over Ireland, Int. J. Cancer, 31, 545–557, 2011. 10750
Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sens. Environ., 111, 519–536, 2007. 10742
Mu, Q., Zhao, M., and Running, S.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115, 1781–1800, 2011. 10743, 10746, 10747
Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J. B., Guo, Z., Jung, M., Maignan, F., McCabe, M. F., Reichle, R., Reichstein, M., Rodell, M., Sheffield, J., Teuling, A. J., Wang, K., Wood, E. F., and Zhang, Y.: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, Geophys. Res. Lett., 38, L06402, doi:10.1029/2010GL046230, 2011. 10742
Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D. G., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K., Wood, E. F., Zhang, Y., and Seneviratne, S. I.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, Hydrol. Earth Syst. Sci., 17, 3707–3720, doi:10.5194/hess-17-3707-2013, 2013. 10743
Murphy, D. and Koop, T.: Review of the vapour pressures of ice and supercooled water for atmospheric applications, Q. J. Roy. Meteorol. Soc., 131, 1539–1565, 2005. 10749
Owe, M., de Jeu, R., and Walker, J.: A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index, IEEE T. Geosci. Remote, 39, 1643–1654, 2001. 10753
Pinty, B., Lavergne, T., Vossbeck, M., Kaminski, T., Aussedat, O., Giering, R., Gobron, N., Taberner, M., Verstraete, M. M., and Widlowski, J.-L.: Retrieving surface parameters for climate models from Moderate Resolution Imaging Spectroradiometer (MODIS)-Multiangle Imaging Spectroradiometer (MISR) albedo products, J. Geophys. Res.-Atmos., 112, D10116, doi:10.1029/2006JD008105, 2007. 10751

HESSD
12, 10739–10787, 2015

Pinty, B., Jung, M., Kaminski, T., Lavergne, T., Mund, M., Plummer, S., Thomas, E., and Widlowski, J.: Evaluation of the JRC-TIP 0.01° products over a mid-latitude deciduous forest site, Remote Sens. Environ., 115, 3567–3581, 2011a. 10751

Pinty, B., Taberner, M., Haemmerle, V., Paradise, S., Vermote, E., Verstraete, M., Gobron, N., and Widlowski, J.-L.: Global-Scale Comparison of MISR and MODIS Land Surface Albedos, J. Climate, 24, 732–749, 2011b. 10751

Priestley, C. and Taylor, R.: On the Assessment of surface heat flux and evaporation using large-scale parameters, Mon. Weather Rev., 100, 81–92, 1972. 10747

Saha, S., Moorthi, S., Pan, H.-L., et al.: The NCEP climate forecast system reanalysis, B. Am. Meteorol. Soc., 91, 1015–1057, 2010. 10754

Seemann, S. W., Borbas, E. E., Knuteson, R. O., Stephenson, G. R., and Huang, H.-L.: Development of a global infrared surface emissivity database for application to clear sky retrievals from multispectral satellite radiance measurements, J. Appl. Meteorol. Clim., 47, 108–123, 2008. 10751, 10755

Simard, M., Pinto, N., Fisher, J., and Baccini, A.: Mapping forest canopy height globally with spaceborne lidar, J. Geophys. Res., 116, G04021, doi:10.1029/2011JG001708, 2011. 10752

Stackhouse, P., Gupta, S., Cox, S., Mikovitz, J., Zhang, T., and Chiacchio, M.: 12-year surface radiation budget data set, GEWEX News, 14, 10–12, 2004. 10751

Su, H., McCabe, M. F., and Wood, E. F.: Modeling evapotranspiration during SMACEX: comparing two approaches for local- and regional-scale prediction, J. Hydrometeorol., 6, 910–922, 2005. 10745

Su, Z.: The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, Hydrol. Earth Syst. Sci., 6, 85–100, doi:10.5194/hess-6-85-2002, 2002. 10743, 10744

Szczypta, C., Calvet, J.-C., Albergel, C., Balsamo, G., Boussetta, S., Carrer, D., Lafont, S., and Meurey, C.: Verification of the new ECMWF ERA-Interim reanalysis over France, Hydrol. Earth Syst. Sci., 15, 647–666, doi:10.5194/hess-15-647-2011, 2011. 10750

Vinukollu, R. K., Sheffield, J., Wood, E. F., Bosilovich, M. G., and Mocko, D.: Multimodel analysis of energy and water fluxes: intercomparisons between operational analyses, a land surface model, and remote sensing, J. Hydrometeorol., 13, 3–26, 2011. 10742

Wang, K. and Dickinson, R.: A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability, Rev. Geophys., 50, RG2005, doi:10.1029/2011RG000373, 2012. 10742

HESSD
12, 10739–10787, 2015

Wang, K., Wan, Z., Wang, P., Sparrow, M., Liu, J., Zhou, X., and Haginoya, S.: Estimation of surface long wave radiation and broadband emissivity using Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature/emissivity products, J. Geophys. Res., 110, D11109, doi:10.1029/2004JD005566, 2005. 10755

Xie, P. and Arkin, P. A.: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, B. Am. Meteorol. Soc., 78, 2539–2558, 1997. 10754

Zhang, K., Kimball, J. S., Nemani, R. R., and Running, S. W.: A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006, Remote Sens. Environ., 46, W09522, doi:10.1029/2009WR008800, 2010. 10742

HESSD
12, 10739–10787, 2015

Table 1. Table summarizing the model inputs. Listed the main inputs, product selected, and original temporal and spatial resolutions, and the satellite sensors used to derive the product.

Variable	Models	Product	Resolution	Sensors
Surf. Radiation	All	SRB	3-hourly/100 km	Several VIS-IR sensors
Surf. temperature polar geostationaty	SEBS	IPMA	 twice-day/1 km hourly/5 km	 AATSR MSG-2, MTSAT, GOES-12
Surf. meteorology temperature humidity wind	 All SEBS/PM/PT SEBS	ERA-Interim	3-hourly/75 km	Assimilation of satellite and other meteo observations
*f*APAR/LAI	SEBS/PM/PT	from ESA GlobAlbedo	8 days / 1 km	VEGETATION, MERIS, MODIS
Soil Moisture	GLEAM	ESA-CCI	daily/25 km	SSM/I, TMI, AMSR-E, ASCAT
Precipitation	GLEAM	CMORPH	30 min/15 km	AMSU-B, AMSR-E, TMI
Snow Water	GLEAM	ESA	daily/1 km	AMSR-E

HESSD
12, 10739–10787, 2015

Table 2. Stations selected to run the models with tower inputs. From left to right the station name; longitude; latitude; Köppen-Geiger Climate Classification (KGCC); International Geosphere-Biosphere International Programme (IGBP) land cover; total number of days with data/no precipitation number of days with data; evaporative fraction for the DJF, MAM, JJA, SON 3-monthly periods.

Name	Lon	Lat	KGCC	IGBP	Days	EF
AU-How	131.15° E	12.49° S	Aw	SV	114/100	0.7/0.7/0.5/0.3
CA-Ojp	104.69° W	53.92° N	Dfc	ENF	126/101	0.2/0.1/0.3/0.5
CA-Qfo	74.34° W	49.69° N	Dfc	ENF	253/166	0.1/0.1/0.4/0.4
DE-Geb	10.91° E	51.1° N	Cfb	CRO	188/113	0.0/0.4/0.5/0.7
DE-Har	7.6° E0	47.93° N	Cfb	MF	105/88	1.0/0.5/0.5/0.7
DE-Kli	13.52° E	50.89° N	Cfb	CRO	275/98	0.0/0.5/0.5/0.0
DE-Meh	13.52° E	50.89° N	Cfb	CRO	444/269	0.0/0.3/0.5/0.5
DE-Wet	11.46° E	50.45° N	Cfb	ENF	384/182	0.1/0.3/0.4/0.6
IT-MBo	11.08° E	46.03° N	Dfb	MF	149/126	0.0/0.6/0.7/0.9
IT-Noe	8.15° E	40.6° N	Csa	WL	182/182	0.7/0.3/0.2/0.3
NL-Ca1	4.93° E	51.97° N	Cfb	CRO	38/22	1.0/0.6/0.6/0.9
PT-Mi2	8.02° W	38.48° N	Csa	SV	275/221	0.5/0.4/0.3/0.4
RU-Fyo	32.92° E	56.46° N	Dfb	MF	374/216	0.0/0.4/0.5/0.4
US-ARM	97.49° W	36.61° N	Cfa	CRO	159/131	0.4/0.5/0.3/0.3
US-Aud	110.51° W	31.59° N	BSk	OSH	219/219	0.5/0.2/0.3/0.5
US-Bkg	96.84° W	44.35° N	Dfa	CRO	174/172	0.6/0.7/0.9/1.0
US-Bo2	88.29° W	40.01° N	Dfa	CRO	192/192	0.3/0.3/0.6/0.3
US-FPe	105.1° W	48.31° N	BSk	GRA	184/184	1.0/0.4/0.6/0.4
US-Goo	89.87° W	34.25° N	Cfa	CRO	183/179	0.7/0.6/0.5/0.6
US-MOz	92.2° W	38.74° N	Cfa	DBF	252/252	0.3/0.4/0.5/0.5
US-SRM	110.87° W	31.82° N	BSk	OSH	139/137	0.2/0.1/0.3/0.4
US-WCr	90.08° W	45.81° N	Dfb	DBF	338/239	0.1/0.2/0.6/0.4
US-Wkg	109.94° W	31.74° N	BSk	GRA	137/137	0.2/0.1/0.2/0.3
US-Wrc	121.95° W	45.82° N	Csb	ENF	146/107	0.4/0.3/0.3/0.5

HESSD
12, 10739–10787, 2015

Table 3. Summary of 24 stations average statistics for 3-hourly and daily tower forcing. EC denotes the model agreement to the evapotranspiration reference from eddy-covariance measurements, and ER is the model agreement to the evapotranspiration reference based on the in situ energy residual. RMSD is given in mm h^{-1} for 3-hourly data (3 h) and in mm day^{-1} for daily data (d).

		R^2		RMSD		MBD [%]	
		EC	ER	EC	ER	EC	ER
PT-JPL	3 h	0.77	0.78	0.08	0.09	53.1	37.4
	d	0.61	0.60	1.04	1.20	47.8	21.8
PM-MOD	3 h	0.58	0.55	0.06	0.12	−6.7	−18.2
	d	0.43	0.41	0.96	1.36	3.8	−11.3
SEBS	3 h	0.64	0.78	0.13	0.10	125.9	78.4
	d	0.45	0.50	1.84	1.68	113.5	48.5
GLEAM	3 h	0.70	0.80	0.08	0.08	31.9	−8.7
	d	0.60	0.66	0.80	0.96	15.6	−14.1

HESSD
12, 10739–10787, 2015

Table 4. Summary of 24 stations average statistics for 3-hourly and daily satellite forcing. ERA denotes the agreement of ERA-Interim evapotranspiration to the in situ reference evapotranspiration. For other abbreviations see Table 3. RMSD is given in mm h^{-1} for 3-hourly data (3 h) and in mm day^{-1} for daily data (d).

		R^2		RMSD		MBD	[%]
		EC	ER	EC	ER	EC	ER
PT-JPL	3 h	0.67	0.68	0.07	0.11	25.8	14.6
	d	0.57	0.49	0.88	1.28	16.5	−1.7
PM-MOD	3 h	0.47	0.47	0.07	0.14	−16.1	−27.2
	d	0.35	0.29	1.04	1.52	−17.9	−34.2
SEBS	3 h	0.59	0.71	0.13	0.11	145.1	148.6
	d	0.42	0.41	2.08	2.00	160.2	123.9
GLEAM	3 h	0.61	0.72	0.08	0.10	22.7	−4.6
	d	0.52	0.52	0.88	1.20	16.2	−10.6
ERA	3 h	0.51	0.45	0.10	0.14	111.3	87.0
	d	0.62	0.50	1.60	1.68	114.7	74.6

HESSD
12, 10739–10787, 2015

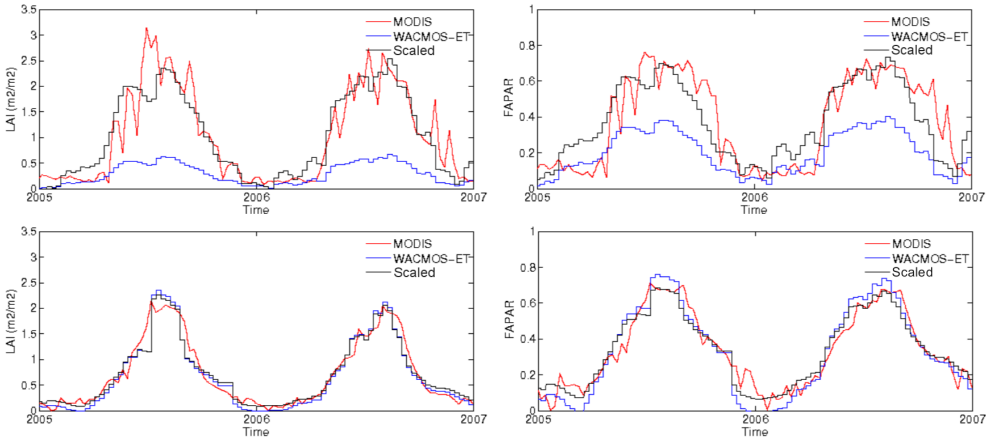

Figure 1. 2005–2006 time series of MODIS MOD15A2 LAI and *f*APAR, WACMOS-ET LAI and *f*APAR, and the MODIS-like LAI and *f*APAR (referred to as *scaled* in the figures) at the tower stations CA-Qfo (top panels) and US-Bkg (bottom panels).

HESSD
12, 10739–10787, 2015

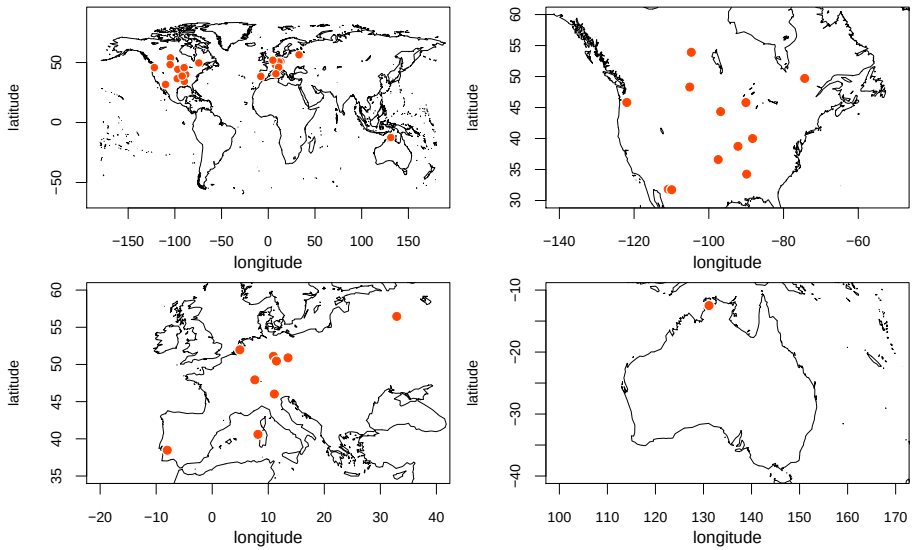

Figure 2. Location of the 24 FLUXNET stations used for the main analysis of the study. They are located on three different continents, encompassing 9 different biomes, i.e. vegetation mosaic, croplands, mixed forests, deciduous forest, savanna, evergreen needleleaf forests, grasslands, woody savanna and shrublands.

HESSD
12, 10739–10787, 2015

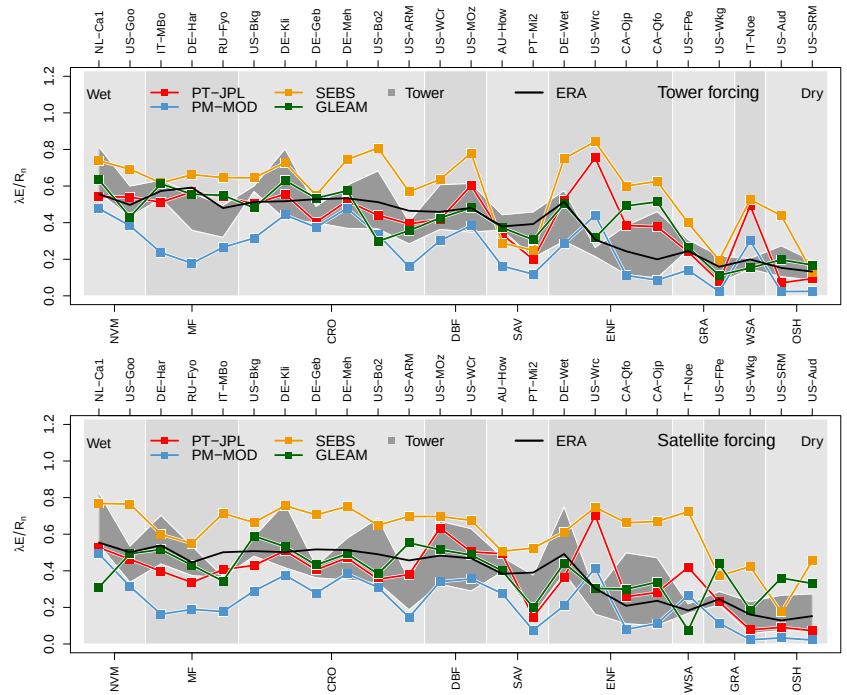

Figure 3. Station means of 3-hourly EC-observed and tower-forced (top panel) and satellite-forced (bottom panel) evaporative fraction against tower reference, as function of biomes, sorted from wet to dry (based on the biome average). The grey area denotes the range of evaporative fraction (grey area) between EC and ER measurements. The black line denotes EF derived from ERA-Interim ET (ERA) and R_n.

HESSD
12, 10739–10787, 2015

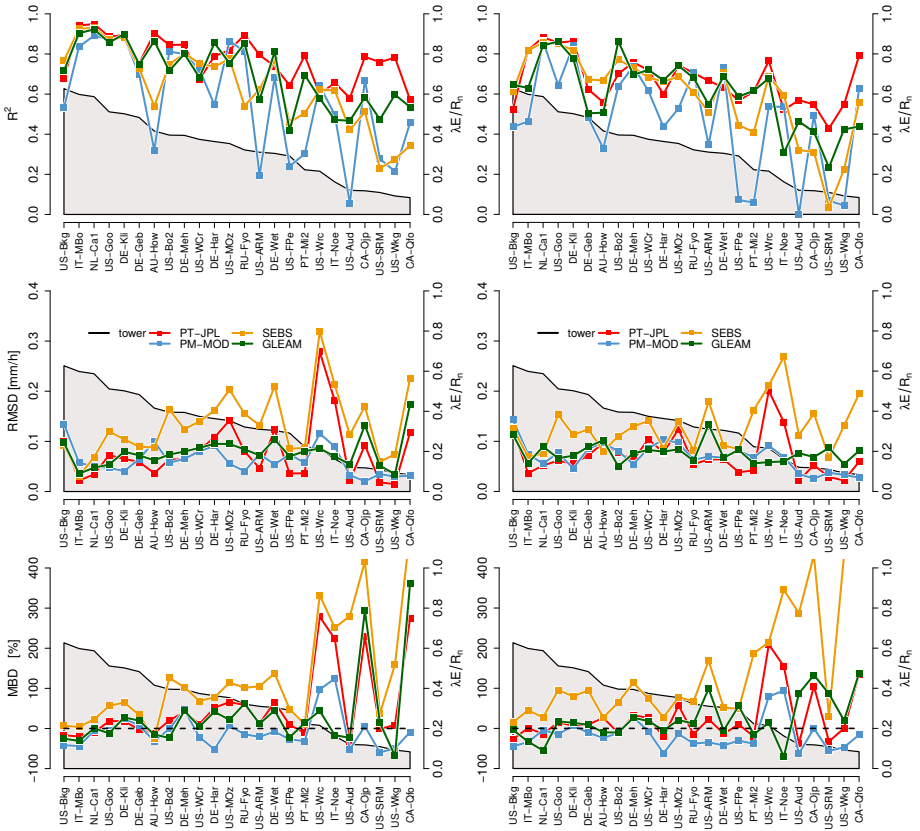

Figure 4. Station mean statistics of 3-hourly model data against EC reference. Left column panels: tower-forced ET, right column panels: satellite-forced ET, top row panels: R^2 correlation coefficient (left *y* axis). Middle row panels: mean bias deviation (MBD, left *y* axis). Bottom row panels: root mean square difference (RMSD, left *y* axis). For all plots the evaporative fraction is given by the grey area (right *y* axis).

HESSD
12, 10739–10787, 2015

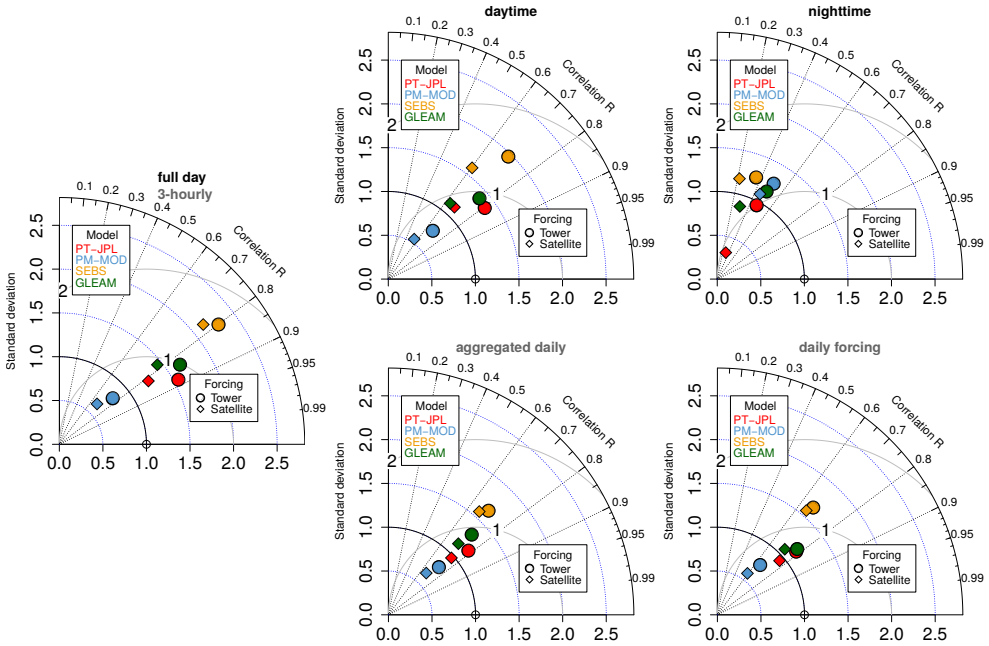

Figure 5. Taylor diagrams of 3-hourly model performance against EC reference in sub-daily periods (top row panels) and as function of temporal resolution (bottom row panels). The left panel shows the average model statistics for full day (compare to top row) and 3-hourly output data (compare to bottom row). Daytime is defined as cases when the cosine of the sun elevation azimuth is > 0.2, nighttime is defined as cases when the cosine of the sun elevation azimuth is < 0.2. Shown are the normalized standard deviation, the normalized RMSD and the correlation coefficient (R).

HESSD
12, 10739–10787, 2015

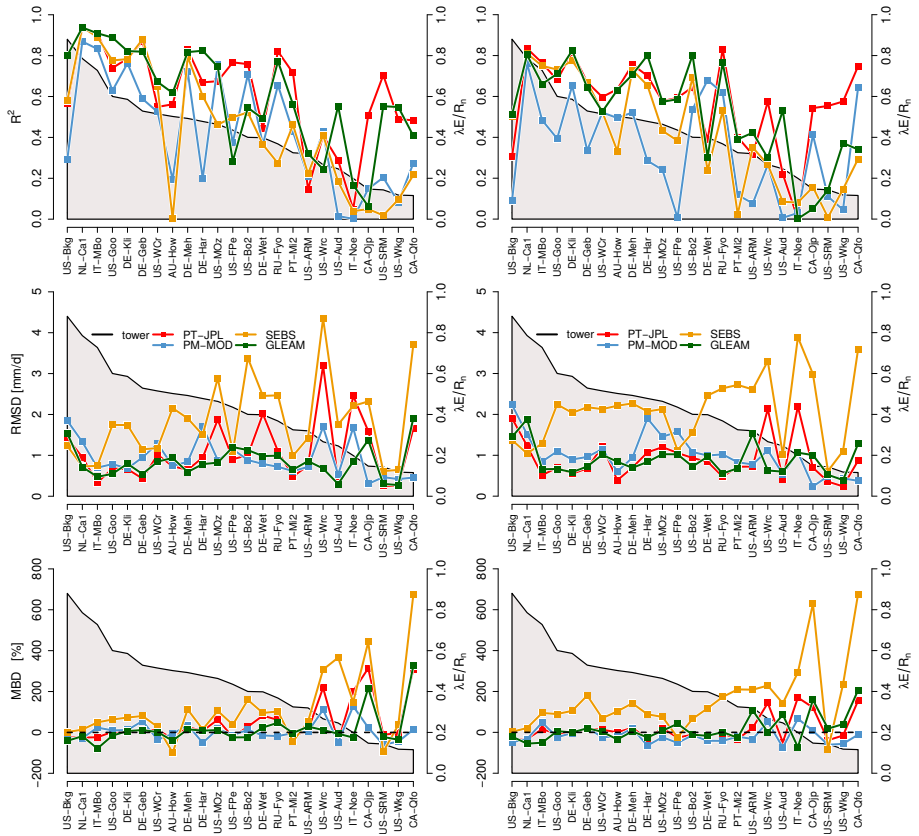

Figure 6. Station mean statistics of daily data from daily input against EC reference. Left column panels: tower-forced ET, right column panels: satellite-forced ET, top row panels: R^2 correlation coefficient (left y axis). Middle row panels: mean bias deviation (MBD, left y axis). Bottom row panels: root mean square difference (RMSD, left y axis). For all plots the evaporative fraction is given by the grey area (right y axis).

HESSD
12, 10739–10787, 2015

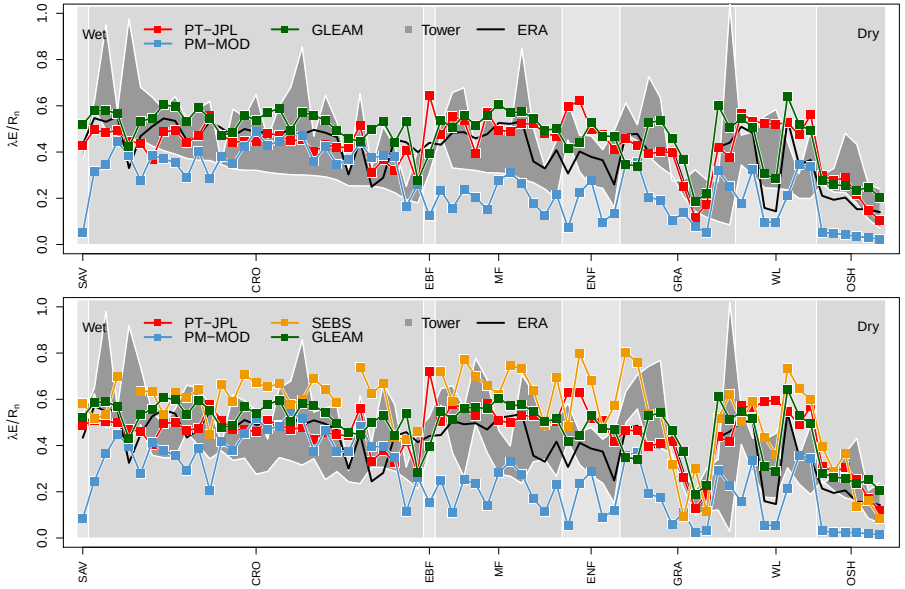

Figure 7. Top panel: station means of 3-hourly sinusoidal gridded satellite-forced evaporative fraction for full days (70 stations) against tower reference, as function of biomes, sorted from wet to dry (based on the biome average). Bottom panel: same as above but for mid-morning only (from 09:00 to 13:00 LT, 67 stations). The grey area denotes the range of evaporative fraction (grey area) between EC and ER measurements. The black line denotes EF derived from ERA-Interim ET and R_n.

HESSD
12, 10739–10787, 2015

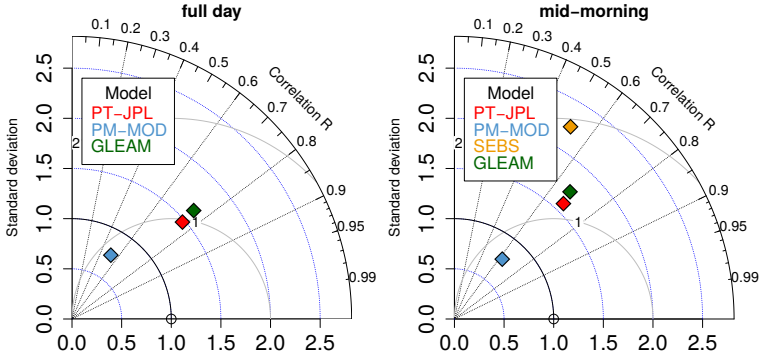

Figure 8. Taylor diagrams plots of sinusoidal gridded model data against tower EC reference. Left panel: full day 3-hourly data compared to 85 stations. Right panel: mid-morning data (from 09:00 and 13:00 LT) compared to 82 stations. Shown are the normalized standard deviation, the normalized RMSD and the correlation coefficient (R).

HESSD
12, 10739–10787, 2015